Notation

Last modified: 2026-09-28 23:45:45 (PDT)

Mathematical notation is not standardized. This section states the conventions these notes use, and the alternatives you may meet in other sources.

Table 1: Notation used in this book

symbol	meaning	LaTeX
$\neg$	not	`\neg`
$\forall$	all	`\forall`
$\exists$	some	`\exists`
$\cup$	union, "or"	`\cup`
$\cap$	intersection, "and"	`\cap`
$\mid$	given, conditional on	`\mid`, `|`
$\sum$	sum	`\sum`
$\prod$	product	`\prod`
μ	mean	`\mu`
E	expectation	`\mathbb{E}`
$x^{\top}$	transpose of x	`x^{\top}`
$'$	transpose or derivative[1]	`'`
$\perp\!\!\!\perp$	independent	`\perp\!\!\!\perp`
$\therefore$	therefore, thus	`\therefore`
η	linear component of a GLM	`\eta`
$\lfloor x \rfloor$	floor of x: largest integer less than or equal to x	`\lfloor x \rfloor`
$\lceil x \rceil$	ceiling of x: smallest integer greater than or equal to x	`\lceil x \rceil`
$\mathbb{1}_A(x)$, $\mathbb{1}(P)$	indicator function (Section): 1 if condition holds, 0 otherwise	`\indic{A}(x)`, `\indicp{P}`

[1] depending on whether it is applied to a matrix or a function

Writing math in Quarto

The third column of Table 1 gives the LaTeX command for each symbol. Quarto and R Markdown documents write math in LaTeX syntax: put inline math between single dollar signs, as in `$x^2$`, and displayed equations between double dollar signs, as in `$$x^2$$`. The equations section of Quarto's Markdown guide shows the details. These notes also use shorthand macros, such as `\floor{x}` for $\lfloor x \rfloor$, defined in the `latex-macros` submodule's `macros.qmd`.

Natural numbers

Exercise 0.1 (Which numbers are natural?)**.** List the elements of the set $\{n \in \mathbb{N} : n < 3\}$. Is your answer the same in every textbook?

Solution

Solution 0.1*.* The answer depends on whether the source counts 0 as a natural number:

- if $\mathbb{N}$ starts at 0, the set is $\{0, 1, 2\}$;
- if $\mathbb{N}$ starts at 1, the set is $\{1, 2\}$.

Both conventions are in common use, so the answer is not the same in every textbook.

Definition 0.1 (Natural numbers (our convention))**.** In these notes, the **natural numbers** are the positive integers:

$$\mathbb{N} \stackrel{\text{def}}{=} \{1, 2, 3, \ldots\}$$

Definition 0.2 (Non-negative integers)**.** The **non-negative integers** are the natural numbers (Definition 0.1) together with 0:

$$\mathbb{N}_0 \stackrel{\text{def}}{=} \{0, 1, 2, 3, \ldots\} = \mathbb{N} \cup \{0\}$$

Example 0.1 (Natural numbers in a data analysis)**.**

- Observation indices start at 1, so we write $i \in \{1, \ldots, n\}$ with $n \in \mathbb{N}$.
- A count outcome, such as the number of hospital visits in a year, can be 0, so its support is $\mathbb{N}_0$, not $\mathbb{N}$.

Other conventions for the natural numbers

Other sources may define $\mathbb{N}$ differently, so check each source's definition before reading its formulas:

- Many sources include 0 in $\mathbb{N}$. The international standard ISO 80000-2 defines $\mathbb{N}$ to include 0, continuing the earlier standard ISO 31-11 (1978).
- Other sources start $\mathbb{N}$ at 1, as these notes do.
- To remove the ambiguity, some sources write $\mathbb{N}_1$ or $\mathbb{Z}^+$ for $\{1, 2, 3, \ldots\}$, and $\mathbb{N}_0$ or $\mathbb{Z}^{0+}$ for $\{0, 1, 2, \ldots\}$.
- The words vary too. "Positive integers" ($\{1, 2, 3, \ldots\}$) and "non-negative integers" ($\{0, 1, 2, \ldots\}$) are unambiguous. "Whole numbers" usually includes 0, but can also mean all of the integers, negative ones included. "Counting numbers" usually starts at 1, but some sources include 0.
- Some older texts write J for the natural numbers.

When a formula's meaning depends on whether 0 is included, we write the set out explicitly, for example $\{0, 1, 2, \ldots\}$ or $\{1, \ldots, n\}$.
Source: Wikipedia, "Natural number", "Terminology and notation" and "Zero as natural number", which cites ISO 80000-2:2019 and the textbooks using each notation.

Percent sign ("%")

The percent sign "%" is just a shorthand for "/100". The word "percent" comes from the Latin "per centum"; "centum" is Latin for 100, so "percent" means "per hundred" (cf. https://en.wikipedia.org/wiki/Percentage)

So, contrary to what you may have learned previously, $10\% = 0.1$ is a true and correct equality, just as $10\text{kg} = 10,000\text{g}$ is true and correct.

Proof

Proof.

$$\begin{aligned} 10\% &= 10/100 \\ &= \frac{10}{100} \\ &= 0.1 \end{aligned}$$

□

You are welcome to switch between decimal and percent notation freely; just make sure you execute it correctly.

Proofs

We can use any of:

- $\therefore$ (`\therefore` in LaTeX),
- $\Rightarrow$ (`\Rightarrow`),
- $\models$ (`\models`)

to denote logical entailments (deductive consequences).

Let's save $\rightarrow$ (`\rightarrow`) for convergence results.

Indicator functions

An **indicator function** is a mathematical function that signals whether an element belongs to a specified set, or whether a given logical condition is satisfied. In statistics and epidemiology, indicator functions are ubiquitous: they represent binary variables, censor and event indicators in survival analysis, membership in subpopulations, and domain restrictions in integrals and sums.

Despite their conceptual simplicity, notation for indicator functions varies substantially across textbooks, research papers, and subfields. This section summarizes the principal notational conventions.

Definition 0.3 (Indicator function)**.** For any subset $A \subseteq \Omega$ of a universal set Ω, the **indicator function** of A is the function $\mathbb{1}_A : \Omega \to \{0, 1\}$ defined by:

$$\mathbb{1}_A(x) \stackrel{\text{def}}{=} \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

More generally, for any logical proposition or predicate P, the **indicator** of P takes the value 1 when P is true and 0 when P is false:

$$\mathbb{1}(P) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } P \text{ is true} \\ 0, & \text{if } P \text{ is false} \end{cases}$$

Example 0.2 (Evaluating set and predicate indicators)**.** Consider the real line $\Omega = \mathbb{R}$, the set of nonnegative numbers $A = [0, \infty)$, and a continuous random variable Y.

1. **Set indicator** $\mathbb{1}_A(x)$:
   - For $x = 3.5$: since $3.5 \in [0, \infty)$, $\mathbb{1}_A(3.5) = 1$.
   - For $x = -2.1$: since $-2.1 \notin [0, \infty)$, $\mathbb{1}_A(-2.1) = 0$.

2. **Predicate indicator** $\mathbb{1}(Y > 5)$:

   - If an observation yields $Y = 7.2$, the predicate $7.2 > 5$ is true, so $\mathbb{1}(7.2 > 5) = 1$.
   - If an observation yields $Y = 4.1$, the predicate $4.1 > 5$ is false, so $\mathbb{1}(4.1 > 5) = 0$.

Two primary notational paradigms: set vs. predicate notation

The vast majority of indicator notations belong to one of two families: **set notation** or **predicate notation**.

Set notation

In **set notation**, the indicator is tied to a set A, which appears as a subscript:

- $\mathbf{1}_A(x)$ (bold numeral one)
- $\mathbb{1}_A(x)$ (blackboard bold numeral one)
- $I_A(x)$ (capital letter I)
- $\mathbb{I}_A(x)$ (blackboard bold letter I)
- $\chi_A(x)$ (Greek letter chi, historically termed the *characteristic function*)

When the function is viewed as a mathematical object in its own right (for instance, as an element of an L^p function space), authors often omit the argument x, writing simply $\mathbf{1}_A$, $\mathbb{1}_A$, or I_A.

Predicate notation

In **predicate notation**, the indicator takes a logical condition, relation, or proposition P directly as its argument or subscript:

- $\mathbb{I}(x \in A)$ or $\mathbb{I}(P)$
- $\mathbf{1}(x \in A)$ or $\mathbf{1}(P)$
- $\mathbb{1}(x \in A)$ or $\mathbb{1}(P)$
- $\mathbb{1}\{x \in A\}$ or $\mathbb{1}\{P\}$
- $I(x \in A)$ or $I(P)$

Predicate notation is especially common in applied statistics and survival analysis, where indicators frequently depend on inequalities involving random variables, such as $\mathbb{1}(T_i \leq t)$ (an event occurring before time t) or $\mathbb{1}(Y_i = 1)$ (a binary outcome).

Equivalence between paradigms

The two paradigms are connected by evaluating the predicate indicator at the membership statement $x \in A$:

$$\mathbf{1}_A(x) = \mathbb{1}(x \in A)$$

Set notation is more natural when the underlying set A has a standard name (such as the support of a distribution or a geometric region). Predicate notation is more natural when the condition involves compound inequalities, such as $\mathbb{1}(0 \leq t \leq u)$.

Iverson bracket notation

In 1962, Kenneth Iverson introduced a compact notation in the programming language APL, later popularized in mathematics and computer science by Donald Knuth: the Iverson bracket.

Definition 0.4 (Iverson bracket)**.** For any logical proposition P, the **Iverson bracket** of P is

$$[P] \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } P \text{ is true} \\ 0, & \text{if } P \text{ is false} \end{cases}$$

The Iverson bracket $[P]$ is the predicate indicator $\mathbb{1}(P)$ (Definition 0.3) in different notation. Under this notation, set membership is written $[x \in A]$.

Example 0.3 (Evaluating Iverson brackets)**.**

- $[3 > 2] = 1$, because $3 > 2$ is true.
- $[2 > 3] = 0$, because $2 > 3$ is false.
- $[4 \in \{1, 2\}] = 0$, because 4 is not an element of $\{1, 2\}$.

Definition 0.5 (Kronecker delta)**.** For integers i and j, the **Kronecker delta** is

$$\delta_{ij} \stackrel{\text{def}}{=} [i = j]$$

that is, $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$ (Definition 0.4).

Example 0.4 (Evaluating the Kronecker delta)**.**

- $\delta_{22} = [2 = 2] = 1$.
- $\delta_{23} = [2 = 3] = 0$.

- The entries of the $p \times p$ identity matrix are Kronecker deltas: $(\mathbf{I}_p)_{ij} = \delta_{ij}$ (see identity matrix).

Strengths and limitations of the Iverson bracket

The primary advantage of the Iverson bracket is algebraic conciseness: it converts domain restrictions in sums and integrals into unrestricted operations. For example:

$$\sum_{x \in A} f(x) = \sum_{x} f(x)[x \in A]$$

With $A = \{2, 4\}$, $f(x) = x$, and x running over $\{1, 2, 3, 4, 5\}$, both sides equal $2 + 4 = 6$.

However, in statistics and epidemiology, square brackets are already heavily overloaded: they denote closed intervals $[a, b]$, conditional expectations $\mathrm{E}[Y \mid X]$, and matrix delimiters. To prevent visual confusion with expectation brackets or intervals, statistical literature predominantly uses $\mathbb{1}$ or I rather than the bare Iverson bracket.

Summary of indicator notations

Table 2 compares the major notations encountered across the literature.

Table 2: Notations for indicator functions across mathematical and statistical literature

Notation style	Typical syntax	Primary fields	Notes and potential ambiguities
Blackboard bold 1	$\mathbb{1}_A(x)$, $\mathbb{1}(P)$	Modern probability, mathematical statistics	Unambiguous; distinct from matrices and scalars; standard in this book.
Bold numeral 1	$\mathbf{1}_A(x)$, $\mathbf{1}(P)$	Probability theory, measure theory	Can be confused with a vector of ones $\mathbf{1} = (1, \dots, 1)^\top$.
Blackboard bold I	$\mathbb{I}(x \in A)$, $\mathbb{I}(P)$	Econometrics, machine learning, statistics	Clear predicate notation; avoids confusion with numerals.
Letter I	$I_A(x)$, $I(P)$	Classical statistics, epidemiology	Can be confused with the identity matrix I or Fisher information $\mathcal{I}$.
Iverson bracket	$[P]$, $[x \in A]$	Computer science, discrete mathematics	Very compact, but square brackets collide with intervals and expectation brackets.

Notation style	Typical syntax	Primary fields	Notes and potential ambiguities
Greek letter χ	$\chi_A(x)$	Real analysis, measure theory	Often termed "characteristic function"; collides with the Fourier transform in probability.

Conventions in this book

In these notes, we standardize on blackboard bold $\mathbb{1}$ via the macros defined in `latex-macros/macros.qmd`:

- `\indic{A}` produces $\mathbb{1}_A$ (set subscript)
- `\indicp{P}` produces $\mathbb{1}(P)$ (predicate in parentheses)
- `\indiccb{P}` produces $\mathbb{1}\{P\}$ (predicate in curly braces)
- `\1{P}` produces 1_P (text numeral with subscript, used in legacy formulas)

Blackboard bold $\mathbb{1}$ is preferred because it avoids all common collisions: it is visually distinct from the scalar 1, the identity matrix I, and the information matrices (I, $\mathcal{I}$).

Key algebraic properties

Indicator functions translate logical operations on events into ordinary arithmetic on real numbers:

For subsets A and B of Ω, with complement $A^c \stackrel{\text{def}}{=} \Omega \setminus A$, and for every $x \in \Omega$:

- **Intersection ("and"):** $\mathbb{1}_{A\cap B}(x) = \mathbb{1}_A(x) \cdot \mathbb{1}_B(x)$
- **Union ("or"):** $\mathbb{1}_{A\cup B}(x) = \mathbb{1}_A(x) + \mathbb{1}_B(x) - \mathbb{1}_A(x) \cdot \mathbb{1}_B(x)$
- **Complement ("not"):** $\mathbb{1}_{A^c}(x) = 1 - \mathbb{1}_A(x)$
- **Idempotence:** $(\mathbb{1}_A(x))^2 = \mathbb{1}_A(x)$
- **Expectation gives probability:** For any event A, the expectation of its indicator is the probability of the event:

$$\mathrm{E}[\mathbb{1}_A] = 0 \cdot \Pr(A^c) + 1 \cdot \Pr(A) = \Pr(A)$$

This fundamental identity connects probability theory directly to linear expectation. It provides the mathematical foundation for empirical proportions, survival curve estimators, and regression models for binary outcomes.

Why is notation in probability and statistics so inconsistent and disorganized?

In grad school, we are asked to learn from increasingly disorganized materials and lectures. Not coincidentally, as the amount of organization decreases, the amount of complexity increases, the amount of difficulty increases, the number of reliable references decreases, and the amount of inconsistency in notation and content increases (both between multiple references and within single references!). In other words, as you approach the cutting-edge of most fields, you start to run into content that hasn't been fully thought through or standardized. This lack of clarity is unfortunate and undesirable, but it is understandable and inevitable.

It's worth noting that calculus was formalized in the 1600s, elementary algebra was formalized around 820, and arithmetic even earlier. And calculus still has several competing notation systems. In contrast, the field of statistics only emerged in the late 1800s and early 1900s, so it's not surprising that the notation and terminology is still developing. Generalized linear models were only formalized in 1972 (Nelder and Wedderburn 1972), which is very recent in terms of the pace of scientific development.

References

Nelder, John Ashworth, and Robert WM Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society Series A: Statistics in Society* 135 (3): 370–84. https://doi.org/10.2307/2344614.